JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2005, **83**, 85–92 NUMBER 1 (JANUARY)

ON SCIENCE AND THE DISCRIMINATIVE LAW OF EFFECT

MICHAEL DAVISON AND JOHN A. NEVIN

UNIVERSITY OF AUCKLAND AND UNIVERSITY OF NEW HAMPSHIRE

This article considers the process of the dissemination of scientific findings from the point of view of the discriminative law of effect. We assume that the purpose of science is to describe the state of the world in an unbiased and accurate manner. We then consider a number of challenges to the unbiased consensual development of science that arise from differences between science that is done, submitted for publication, and published. These challenges arise from the differential reinforcers for both research and publication delivered by journals and editors for novel results, the undervaluation of systematic replication and findings of invariance, and general lack of reinforcers for failed replications. All these challenges bias science toward searching for, reporting, and valuing novel results and consequently lead to a biased and erroneous view of the world. We suggest that science should be approached more conservatively, and that a reevaluation of the value of replication, and especially failed replication, is in order.

Key words: replication, discriminative law of effect

After Keeton (1971) reported that navigation by homing pigeons carrying magnetized bars was disrupted relative to pigeons carrying brass bars, a number of investigators attempted to demonstrate pigeons' sensitivity to magnetic fields in laboratory settings. This was achieved by Bookman (1977) using pigeons in a flyway. But in 1987, after informal discussions among interested researchers suggested widespread failures to replicate the effect systematically, four of the negative findings were assembled and published in a special section of *Animal Learning & Behavior.* In the introduction, Griffin (1987) remarked that ''Many other experiments of this general kind have led to negative results, which are not ordinarily published'' (p. 108). This story illustrates both the importance of systematic replication in science and the fact that negative results may be published slowly (or not at all). Here, we discuss the process of replication and the role of negative results from the perspective of the discriminative law of effect (Davison & Nevin, 1999).[1]

Reprints may be obtained from either author (e-mail: m.davison@auckland.ac.nz or jnevin@cisunix.unh.edu).

[1] Recently, discrimination of magnetic fields has been demonstrated in a laboratory analog to a signal-detection procedure (Mora et al., 2004) in which pigeons had to walk through magnetic fields. It seems likely that movement of the pigeon through the field, rather than movement of the field through the pigeon, is critical for magnetic discrimination.

THE DISCRIMINATIVE LAW OF EFFECT

The canonical paradigm that we have used to understand stimulus and reinforcer control derives from signal-detection research. In its simplest form, there are two states of the world (e.g., stimulus present and stimulus absent) and two available responses (e.g., Yes, No). Saying ''Yes'' when the stimulus is present, and ''No'' when it is absent, are correct responses from the purview of the experimenter and provide occasional reinforcers. The other two stimulus-response combinations (''Yes'' when the stimulus is absent: false alarms; and ''No'' when the stimulus is present: misses) are errors, and are not followed by reinforcers. The omni-scient experimenter (or the equipment) knows which stimulus was presented and which response was emitted, and therefore can provide reinforcing or punishing consequences with no uncertainty. The subject may be less clear about which response produced which consequence in the presence of which stimulus. Specifically, if the subject confuses the stimuli or the responses that lead to a reinforcer, the reinforcer may strengthen a stimulus-response pair other than the objective, experimenter-defined pair. The discriminative law of effect is a quantitative theory that measures the difference between the experimenter's and the subject's knowledge of the world and its contingencies. In this model, the relation between the stimuli and the correct responses

	Yes	No
Stimulus	50	0
No Stimulus	0	50

Fig. 1. The experimenter's knowledge of the contingencies arranged in a conditional-discrimination experiment.

	Yes	No
Stimulus	50.25	7.5
No Stimulus	7.5	50.25

Fig. 2. The state of the objective world represented in Figure 1 according to the subject, assuming imperfect stimulus-response discriminability ($d_{sb} = 10$) and imperfect response-reinforcer discriminability ($d_{br} = 20$). We assume equal frequencies of reinforcers for the Yes and No responses.

is characterized by a parameter called d_{sb} and the relation between the responses and their consequences is characterized by a parameter called d_{br}. To the extent that d_{sb} and d_{br} are less than infinite, errors are inevitable.

For example, in a 100-trial experiment with stimuli presented on a random half of the trials and feedback or reinforcement for all correct responses, the experimenter's knowledge of the world may be represented in Figure 1. According to Davison and Nevin (1999), the subject's knowledge of the world will approximate the experimenter's knowledge of the world only to the extent that the subject is able to discriminate perfectly the contingencies embodied in Figure 1. If the discrimination is less than perfect, then some of the reinforcers delivered for one response in the presence of one stimulus may appear to the subject to have arisen from making that response in the presence of a different stimulus (a stimulus-behavior error), or from making a different response in the presence of that stimulus (a behavior-reinforcer error), or from both of these errors. If these discrimination failures occur, the experimenter-delivered reinforcers will appear to generalize to other cells of the matrix of Figure 1. The parameter that measures stimulus-response discriminability is d_{sb}, and the parameter that measures response-reinforcer discriminability is d_{br}. Their values range from 1 (no discrimination) to infinity (perfect discrimination). Thus if $d_{sb} = 10$ (about one stimulus-behavior error in 10 trials) and $d_{br} = 20$ (one error in 20 trials), the subject will experience some reinforcers that apparently follow experimenter-defined errors (No|Stimulus and Yes|No stimulus). For example, the apparent reinforcers for saying No given a stimulus presentation are:

$$\text{No}|\text{Stimulus} = \frac{\text{Yes}|\text{Stimulus}}{d_{br}} + \text{No}|\text{Stimulus} + \frac{\text{Yes}|\text{No Stimulus}}{d_{sb}d_{br}} + \frac{\text{No}|\text{No Stimulus}}{d_{sb}} = \frac{50}{20} + 0 + 0 + \frac{50}{10} = 7.5.$$

Similar calculations are carried out for all cells of the matrix in Figure 1. Thus, according to the discriminative law of effect, the subject's knowledge of the world will be as shown in Figure 2.

In order to estimate the values of d_{sb} and d_{br}, or other theoretical parameters such as d' in signal detection theory, the probability of Yes|No stimulus must be estimated at least as accurately as Yes|Stimulus.

Our example in Figure 2 assumes equal feedback or reinforcement for the two kinds of correct responses. However, if correctly saying Yes leads to larger or more frequent reinforcers than for correctly saying No, responding is biased toward Yes even if the quality of the subject's information about the world (i.e., d_{sb} and d_{br}) is unchanged. And if the bias is substantial, measured accuracy may decrease. The effect of a five-fold difference in reinforcer value on the subject's payoff matrix is shown in Figure 3. To calculate this matrix, we assumed 100 reinforcers for Yes|Stimulus and 20 for No|No stimulus, with the same values of d_{sb} and d_{br} as we used to calculate Figure 2.

The relation between this paradigm and the scientific paradigm can readily be seen. The subject becomes the scientist, the stimuli are the states of the world, and the responses are

	Yes	No
Stimulus	100.1	7
No Stimulus	11	20.5

Fig. 3. The state of the objective world represented in Figure 1 according to the subject, assuming imperfect stimulus-response discriminability (d_{sb} = 10) and imperfect response-reinforcer discriminability (d_{br} = 20). Here, we assume five times more reinforcers for Yes responses than for No responses.

	Yes	No
Effect?	111.1	27.5

Fig. 4. Because the objective state of the world is unknown (that is, we cannot know whether the world is presenting a stimulus or no stimulus), there can be no differential reinforcement for the scientist with respect to the state of the world, and hence for the scientist's decision to submit for publication (Yes). Thus the biased matrix of Figure 3 collapses into a single line showing a strong bias for Yes.

reports of the presence or the absence of an effect. Before beginning an experiment, the scientist is unsure about the state of the world but, as we shall argue below, the payoffs for reporting a novel effect are larger than those for reporting no effect, as in Figure 3. However, the big difference between the conditional-discrimination and the scientific paradigm is that in science there is no uber-experimenter (or equipment) that knows the state of the world. Therefore, differential reinforcement cannot be made contingent on correspondence between a scientist's conclusions and the state of the world, and the biased matrix collapses to a single row as shown in Figure 4.

We argue below that the bias toward reporting an effect over reporting a noneffect suggested in Figure 4 is characteristic of empirical research that is submitted for publication. But, before the results are generally available to the scientific audience, there is another Yes–No decision that must be made: to publish or not. Clearly, an editor has no more knowledge of the state of the world than the researcher, so the decision to publish must be based on other considerations that we discuss below. An editor's matrix is shown in Figure 5, assuming the same total reinforcement as in Figures 2 to 4 and with a further five-fold difference in reinforcer value for publication of new positive findings. The overall effect of these successive submission and editorial biases is to establish a very strong bias toward publication of positive results in the scientific literature, even though the process is supposed to be neutral with respect to the state of the world where positive and negative findings may be equally likely.

SCIENCE—SEEN AND UNSEEN

Let us first distinguish between science that is done and science that is seen to be done (that is, submitted for publication)—often called the "file-drawer problem." It has been said (e.g., by S. S. Stevens to the second author) that no research project is complete until it is written up and submitted to a journal for scrutiny by reviewers. But not all science that is done is seen by reviewers. For example, student projects (sometimes designed to replicate or extend a published report) are unlikely to be seen outside the professor's lab unless something quite striking emerges. In addition, busy investigators are more likely to submit their most exciting findings than routine replications. Thus there is no way to estimate the amount or the quality of unseen research (although some idea of the former, and perhaps the latter, can be gained from conference presentations and barroom talk). Finally, if the quantity and quality of unseen research is unknown, then so is its contribution to the development of science—perhaps most of the unseen research supports a reported finding, but perhaps it does not. Thus what an editor sees is a biased sample of research that has been conducted by investigators.

SELECTION FOR AGREEMENT

Because no result can be validated against a state of the world, there can be no ongoing

	Yes	No
Publish?	132.1	6.5

Fig. 5. The reinforcer matrix supporting an editor's decision whether or not to publish a result. The true state of the world remains unknown, and further biases operate (we assume another five-fold bias) for publishing novel positive findings.

training in science that says ''Yes, you got it right'' or ''No, you got it wrong.'' At best, your result can agree with that of another researcher. In this respect, the scientist is like the quality-control inspector (a pigeon) discussed by Verhave (1966). Verhave showed that pigeons could be trained to make fine discriminations between defective and intact pills on a conveyor when the experimenter occasionally provided a known defective pill and explicit reinforcement for responses to it. But Verhave thought this process inefficient, so he instituted a procedure in which 2 pigeons viewed the same pill, and reinforcers were provided occasionally if the pigeons agreed. If pigeons were initially trained individually to a high level of accuracy and then put on the agreement regime, the fine discrimination could be maintained indefinitely with no known defective pills put in by the experimenter—because the probability of 2 pigeons both agreeing and being wrong on any given trial is very small indeed. Further, if a discrimination-naive pigeon is placed in an agreement procedure with a trained pigeon, the naive pigeon very soon gets trained up to show high discrimination levels (Cooper, 1972). This is very like the process of science. Young scientists are trained by practiced scientists and then thrown into the situation in which their success is determined by agreeing with the findings of other scientists, or by reporting novel results that others will agree with later. For science as a whole, this process *should* establish and maintain an accurate consensus judgment of the state of the world. But notice the bootstrapping process: With reinforcement for agreement, if the trainer pigeon has a high rate of errors, the trainee also will come to have a high error rate.

If we assume that the real state of the world is that a stimulus (effect) is present, and that each experiment is of sufficient quality (high enough d_{sb}, but still with error possible) to detect a state of the world, then over a series of independent experiments there should be more Yes findings than No findings. Therefore, reinforcers accumulate in the Yes cell of Figure 4 and the effect will be well-supported in the sense that the behavior of scientists will be strongly biased toward the effect being present, and the result becomes enshrined in research reviews and textbooks. Equally, if an effect was not present in the world, ''No'' findings should predominate, the behavior of scientists will become progressively biased toward ''No,'' and the purported effect (e.g., transfer of learning in planaria by cannibalism) will be dismissed.[2] However, if a series of experiments are not independent (perhaps via reinforcement for agreement), reinforcers can potentially accumulate in the erroneous ''Yes'' cell and an effect that does not exist can become well-supported.

The above is no more than a behavioral restatement of the widely held notion that science moves inexorably toward a true depiction of the world through the replication process—that is, that true positive results will be repeated and false positives will prove to be unrepeatable. But even if all competently conducted research was written up and submitted for publication, there are two interacting problems that need to be confronted. The first is publication policy, and the second is publication salience.

PUBLICATION POLICY

Clearly, the scientific literature represents all science only if all research, or a purely random sample of all research, is published. But the scientific literature does not, of course, for a number of good reasons. Among these reasons are quite reasonable questions of the quality of research, the analysis of the research data, the fact that a result already may have been reported and replicated sufficiently, and the publication policies of journals.

Journal publication policies vary considerably, from the exclusive publication of sexy new results right through to the publication of any result given that the page charge is met. This dimension appears to be correlated with the Impact Factor measure, and is probably quite strongly correlated with the differential reinforcement (in terms of frequency

[2] After McConnell's (1962) initial finding of transfer by cannibalism circulated through the worm-runner community, there may have been many failed replications (including one by the second author in 1964) that were not submitted or published. A well-controlled multigroup study by Hartry, Keith-Lee, and Morton (1964) repeated McConnell's effect and showed that the same effect was obtained by cannibalizing untrained worms (conditional stimulus only and handling controls). A literature search (PsycINFO) found that the last study on this topic was published in 1967.

and magnitude of reinforcers) of ''Yes'' responses. Thus science will progress fast via the high-impact journals if (and only if) the published results accurately reflect the state of the world and are not errors (false alarms). If they are the latter, a serious problem ensues: Eliminating such errors (with their differential reinforcers) will be difficult, and may be achieved only slowly in journals of considerably lesser impact that have little effect on scientists' behavior.

From our experience, to publish a failed replication at the same journal level as the original findings requires that the replication be *much* more extensive (i.e., more conditions, multiple subexperiments, etc.) than the original (see, for example, Charman & Davison, 1982, on the ''short-component effect'' in multiple schedules). This is not the kind of research that can be given to a graduate student to complete, and indeed graduate students generally look askance at replications, and particularly at failed replications. Such research, they believe, is far from being sexy, and may well lead to lower grades and less enthusiastic job recommendations than new, astounding findings. Moreover, failed replications can lead to residual doubts in the scientific community about a researcher's technique. As with any mass media, graders and scientists love the unusual, and journals value the added impact that such research gives them. All these factors contribute to the publication bias suggested in Figure 5.

Replication, particularly direct replication that fails, is the poor relation of science. In one sense, this is right and proper—failures may often result from poor design and/or technique, and indeed these may be hard to detect. And it is widely recognized that the ability to detect the true absence of an effect requires substantial statistical power. Nevertheless, the valuation of failed replications generalizes quite strongly, and unfairly, to an implication that any failure to replicate probably is a quality failure, the view often being that although the research looks right, there must have been something odd somewhere. As a result, such results either do not get published at all, or get published in low-status journals where they have little effect (indeed, the very publication of the research in these outlets supports the notion that the research must be poor because it was probably rejected by a series of better journals before ending up there). This is double jeopardy. It results in a publication bias toward ''Yes'' that can easily be maintained when the state of the world is ''No.'' This is a Type-1 error in statistics.

PUBLICATION SALIENCE

The process of science is cumulative, and historical results continue to be available now because they were published in journals that are available now. Although a prior result can be downgraded in value by advances in equipment and analysis, there is no memory effect in science—it is not the case that previous results have diminished value simply because of the passage of time. Therefore, the scientific consensus should be based on an evenhanded consideration of all published findings. But scientists are organisms, and some findings are more salient than others, in part because of the journals in which they appear and in part because of citation frequency. For example, even if published, null results may not be widely cited, whereas positive findings are likely to retain their salience through repeated citation.[3]

In addition, there are salience effects based on primacy. The development of a new line of research starts with agreement that the state of the world is unknown (though there may well be an existing bias toward expecting one result or another based on previous data and theory). The first *published* research in an area has a privileged position because it adds a particular value to scientists' view of the world. Subsequent *published* research either adds in value to this view, or decreases it. A failed replication should return the state of scientists' behavior to what it was before the first report—perhaps still biased by existing theory, although this bias, too, should have been decremented. We discussed above the difficulty of publishing failed replications in

[3] A notable, and highly cited exception here is the finding that amnesics do not differ from controls on word completion, whereas they do differ from controls on free- and cued-recall and recognition tasks (Graf, Squire & Mandler, 1984). This finding was salient, however, because it occurred in the historical context of confusion about the ways in which memory in amnesics was impaired, and perhaps a general expectation that amnesics should be impaired on all memory tasks.

equally prestigious journals, or even at all. But even if published, a failed replication cannot acquire the salience of being the first publication because there is no prior finding to replicate. Thus a failed replication has less impact than the initial positive finding, much as a successful replication has less value than the original research.

In addition to their diminished salience, null results seem to reside in abeyance, rather than in the mainstream of science, because of the statistical view that allows a null hypothesis only to be rejected, not accepted. The presumption is that if the person reporting the failed replication had used better technique, more conditions and subjects, a more appropriate analysis, the null hypothesis could have been rejected. This is a major problem if your metascientific view is, like ours, that one job of science is to find and to mine invariances—areas in which things stay constant and over which a particular theory can be applied successfully. This conservative view should bias our behavior towards ''No'' before we commence an experiment.

INVARIANCE

Within science, as Nevin (1984) has discussed, a finding of invariance has a special place. A very important invariance comes from direct and systematic replication of a nominal finding, but equally important is the direct or especially the systematic replication of the absence of a finding. Knowing that the variation of an independent variable does not have an effect is important in allowing the scientist to assert independence from that variable, and to do so generally if the independent variable is changed over a wide range.

We take as an example the generalized matching law for reinforcer rate and magnitude in concurrent schedules, which is usually written as:

$$\log \frac{B_1}{B_2} = a \log \frac{r_1}{r_2} + b \log \frac{m_1}{m_2} + \log c,$$

where B_1 and B_2 are response rates, r_1 and r_2 are reinforcer rates, m_1 and m_2 are reinforcer magnitudes, a and b represent the sensitivity of response allocation to reinforcer rate and magnitude, respectively, and c represents inherent bias. McSweeney (1975) varied pigeons' body weights and total reinforcement in concurrent variable-interval (VI) schedules with a 4:1 ratio of reinforcer rates. Her data show matching to relative reinforcer rates ($a = 1$) with no consistent effect on log response ratios. Therefore, we do not need a parameter for deprivation in Equation 1. (This, of course, is a rather limited invariance: Preference between different foods, or between food and water, is affected by the deprivation for food or water.) McSweeney's results also suggested that a was invariant with respect to overall reinforcer rate, which was later shown to be incorrect by Alsop and Elliffe (1988), Elliffe and Alsop (1996), and Logue and Chavarro (1987). In related research, Davison (1988) found that sensitivity to reinforcer magnitude depended inversely on reinforcer rate, but McLean and Blampied (2001) showed that the inverse was not true: sensitivity to relative reinforcer frequency is unaffected by variations in absolute reinforcer magnitude. Thus data on the question of whether choice sensitivity depends on absolute values of reinforcer rate or magnitude, and whether sensitivity to rate and magnitude ratios are invariant with respect to each other, are mixed, and we need both systematic and direct replications that should be published even—or especially—if they report failures to confirm earlier data.

THE EDITORIAL PROCESS

Editors do not know the state of the world, but their behavior is biased toward certain results before the review process begins. They know, we assume, existing data and theories, and some theories may be dear to them. Their behavior also may be biased toward maintaining or enhancing the journal's impact factor. With these existing biases (and we always use the term ''bias'' in the technical, rather than the pejorative sense), their jobs are to decide how well the submitted research was able to detect the state of the world. However, there are no reinforcers directly available for this discrimination, except in the sense that subsequent published research could (in theory) support their decision. But we have already discussed the sources of bias in this subsequent process. Indeed, it may be hard for a particular editor to accept, in quick succession, two equal-quality papers that come to radically different conclusions, especially if the first one found, in a theory-free area, an effect of a variable. Additionally, editors seem loath to accept papers in theory-free areas, often requiring for the paper to be published that some prior theoretical or data reason for doing the research is found and displayed. How important is a paper that demonstrates that a theory which is silent on the effect of a particular independent variable is correct in its silence? We think this important, but others may not.

WHAT TO DO?

The processes we have described above would lead inexorably and dynamically to an accurate scientific description of the world if there were no editorial, reviewer, and other biases toward publishing novel effects and differences, and away from publishing null results and failures to replicate. But the biases that we have discussed lead to an increasing number of false alarms (Yes|No stimulus) because of the strong biases to publishing and valuing highly results of the "Yes" variety, and not publishing, and valuing lower, results of the "No" variety. As a result, much of psychology (and maybe of science generally) may well be a Type-1 error. Theory-building in such an environment is fraught with difficulties especially if, as is most often the case, editors require that research reports are placed in the context of theories that have a wide domain which cover all existing results bearing on a theory. It does not look good, and is generally unconvincing to editors and readers alike, if theoreticians try to argue away particular findings as being errors. It seems that we assume, foolishly, that we have an accurate description of the world, whereas in reality we have a biased description of the world.

CONCLUSION

Doing science is consensual behavior, and we have argued that it is biased consensual behavior because of publication policies which themselves are biased by reinforcers that should be extraneous to science. We believe that understanding the sources of bias, and redressing some of the current imbalances by actively encouraging submission of failures to replicate, or to find an invariance—in conjunction with a single-subject approach—could lead to a science that much better represents the world and how it works. We are aware that our suggestions would likely decrease the status and impact factor of a journal that followed them—but in the long term, this latter consideration is irrelevant.

We close with an observation based on the discriminative law of effect. In our analyses of the effects of reinforcement for errors in a detection task, we found that d_{sb} and d_{br} were invariant even though measured discrimination was sharply reduced (Davison & Nevin, 1999). In scientific research, d_{sb} and d_{br} correspond roughly to the true magnitude of an effect and its discriminability by a given experimental method. Thus publication of erroneous results need not reduce the potential for continuing research to describe the state of the world as accurately as possible.

REFERENCES

Alsop, B., & Elliffe, D. (1988). Concurrent-schedule performance: Effects of relative and overall reinforcer rate. *Journal of the Experimental Analysis of Behavior, 49,* 21–36.

Bookman, M. A. (1977). Sensitivity of the homing pigeon to an earth-strength magnetic field. *Nature, 267,* 340–342.

Charman, L., & Davison, M. (1982). On the effects of component durations and component reinforcement rates in multiple schedules. *Journal of the Experimental Analysis of Behavior, 37,* 417–439.

Cooper, D. M. (1972). *Discrimination performance by more-than-one subject.* Unpublished master's thesis, The University of Auckland, Auckland, New Zealand.

Davison, M. (1988). Concurrent schedules: Interaction of reinforcer frequency and reinforcer magnitude. *Journal of the Experimental Analysis of Behavior, 49,* 339–349.

Davison, M., & Nevin, J. A. (1999). Stimuli, reinforcers, and behavior: An integration. *Journal of the Experimental Analysis of Behavior, 71,* 439–482.

Elliffe, D., & Alsop, B. (1996). Concurrent choice: Effects of overall reinforcer rate and the temporal distribution of reinforcers. *Journal of the Experimental Analysis of Behavior, 65,* 445–463.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 164–178.

Griffin, D. R. (1987). Foreword to papers on magnetic sensitivity in birds. *Animal Learning & Behavior, 15,* 108–109.

Hartry, A. L., Keith-Lee, P., & Morton, W. D. (1964). Planaria: Memory transfer through cannibalism reexamined. *Science, 146,* 274–275.

Keeton, W. T. (1971). Magnets interfere with pigeon homing. *Proceedings of the National Academy of Sciences of the United States of America, 68,* 102–106.

Logue, A. W., & Chavarro, A. (1987). Effects on choice of absolute and relative values of reinforcer delay, amount, and frequency. *Journal of Experimental Psychology: Animal Behavior Processes, 13,* 280–291.

McConnell, J. V. (1962). Memory transfer through cannibalism in planarians. *Journal of Neuropsychiatry, 3,* 542–548.

McLean, A. P., & Blampied, N. M. (2001). Sensitivity to relative reinforcer rate in concurrent schedules: Independence from relative and absolute reinforcer duration. *Journal of the Experimental Analysis of Behavior, 75,* 25–42.

McSweeney, F. K. (1975). Concurrent schedule responding as a function of body weight. *Animal Learning & Behavior, 3,* 264–270.

Mora, C. V., Davison, M., Wild, M. J., & Walker, M. M. (2004). Conditioning analysis of magnetoreception and its mechanism in the homing pigeon (*Columba livia*). *Nature, 432,* 508–511.

Nevin, J. A. (1984). Quantitative analysis. *Journal of the Experimental Analysis of Behavior, 42,* 421–434.

Verhave, T. (1966). The pigeon as a quality control inspector. *American Psychologist, 21,* 109–115.

Received April 7, 2004
Final acceptance September 25, 2004